IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

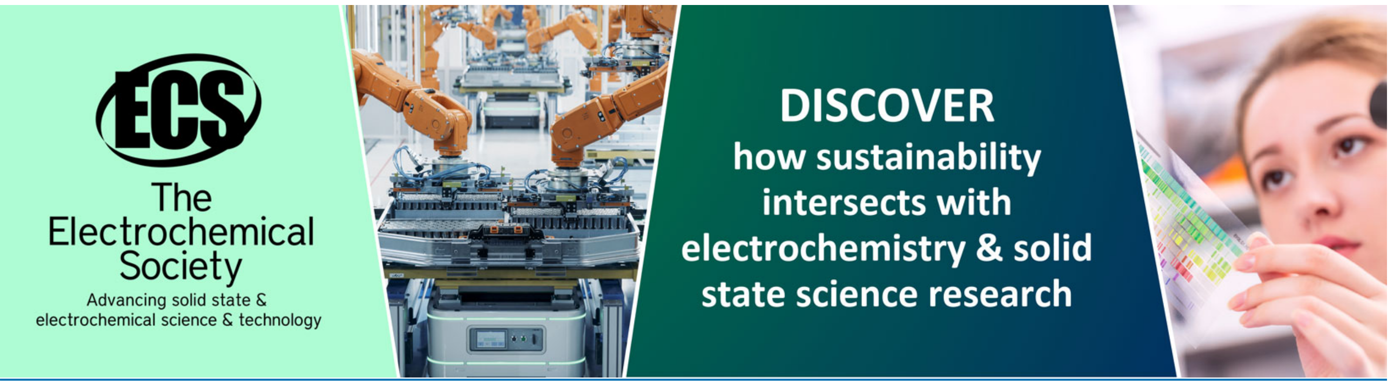
In-store Customer Shopping Behavior Analysis by Utilizing RFID-enabled Shelf and Multilayer Perceptron Model

To cite this article: Ganjar Alfian *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **803** 012022

View the article online for updates and enhancements.

You may also like

- Customer churn factors detection in Indonesian postpaid telecommunication services as a supporting medium for preventing waste of IT components
B Saputro, S Ma'mun, I Budi et al.
- Cosmetics Customer Segmentation and Profile in Indonesia Using Clustering and Classification Algorithm
Sari Hartini, Windu Gata, Sigit Kurniawan et al.
- Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes
Y Yulianti and A Saifudin



This content was downloaded from IP address 18.222.111.24 on 06/05/2024 at 09:06

IOP Conf. Series: Materials Science and Engineering **803** (2020) 012022 doi:10.1088/1757-899X/803/1/012022

In-store Customer Shopping Behavior Analysis by Utilizing RFID-enabled Shelf and Multilayer Perceptron Model

Ganjar Alfian[1], Muhammad Syafrudin[2], Jongtae Rhee[2], Pavel Stasa[3], Agus Mulyanto[4], Agung Fatwanto[4]

[1] Nano Information Technology Academy, Dongguk University, Seoul, Korea
[2] Department of Industrial and Systems Engineering, Dongguk University, Seoul, Korea
[3] Department of Economics and Control Systems, Faculty of Mining and Geology, VSB–Technical University of Ostrava, Czech Republic
[4] Teknik Informatika, Universitas Islam Negeri Sunan Kalijaga, Indonesia

Email: ganjar@dongguk.edu, udin@dongguk.edu, jtrhee@dongguk.edu, pavel.stasa@vsb.cz, agus.mulyanto@uin-suka.ac.id, agung.fatwanto@uin-suka.ac.id

Abstract. Understanding customer shopping behavior in retail store is important to improve the customers' relationship with the retailer, which can help to lift the revenue of the business. However, compared to online store, the customer browsing activities in the retail store is difficult to be analysed. Therefore, in this study the customer shopping behavior analysis (i.e., browsing activity) in retail store by utilizing radio frequency identification (RFID)-enabled shelf and machine learning model is proposed. First, the RFID technology is installed in the store shelf to monitor the movement tagged products. The dataset was gathered from receive signal strength (RSS) of the tags for different customer behavior scenario. The statistical features were extracted from RSS of tags. Finally, machine learning models were utilized to classify different customer shopping activities. The experiment result showed that the proposed model based on Multilayer Perceptron (MLP) outperformed other models by as much as 97.00%, 96.67%, 97.50%, and 96.57% for accuracy, precision, recall, and f-score, respectively. The proposed model can help the managers better understand what products customer interested in, so that can be utilized for product placement, promotion as well as relevant product recommendations to the customers.

1. Introduction

Understanding customer shopping behavior will provide useful input for the managers to improve the effectiveness of marketing and service quality. For online store, customer shopping behavior can be analyzed based on click streams (pages visited) and customer shopping carts (purchased products) data [1]. However, retail stores still lack effective method to comprehensively identify shopping behaviors. The retail stores only provided sales history data, which difficult to understand the customer behavior before they leave the store, such as what products the customer browse or interested in. Therefore, it is necessary to provide a solution to understand customer behavior in retail store.

Radio frequency identification (RFID) is well-known auto identification technology that has been applied to many areas especially for item-level in the retail store. Athauda et al [2] proposed RFID based smart shopping trolley. The UHF antenna is mounted to trolley so that tagged products can be traced in real-time. Benes et al [3] utilized RFID to detect customer movement. By utilizing tagged cart, it is possible to track movement and the time spent of customers. Syaekhoni et al [4] utilized RFID for

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

IOP Conf. Series: Materials Science and Engineering **803** (2020) 012022 doi:10.1088/1757-899X/803/1/012022

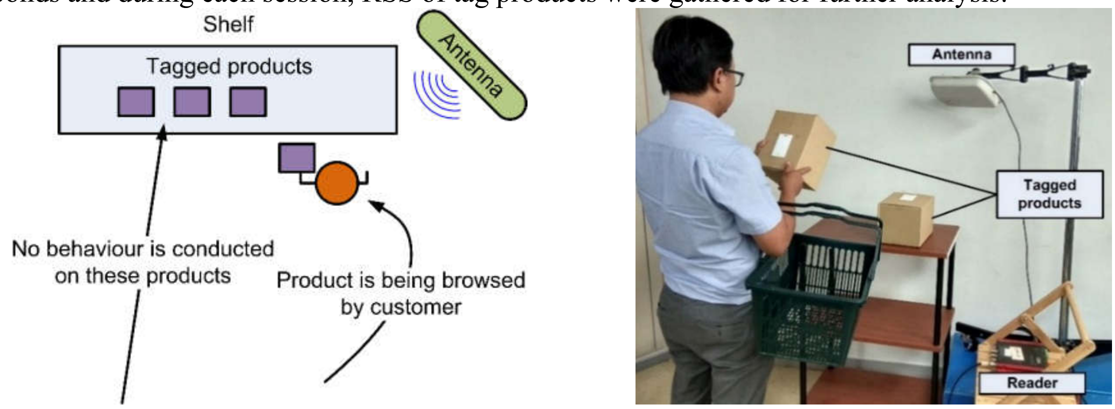
analyzing shopping path of customers in retail store. By combining shopping path with purchasing data, the clustering analysis can be applied to group the customers and uncover the interesting characteristics of each group. Regarding customer browsing detection, Choi et al [5] and Zhou et al [6] utilized RFID to understand several customer activities such as picking up, browsing, returning, interest in and match up. They utilized count or total time of being read for tag [5] and phase readings [6] as input to distinguish different type of customer behavior.

Furthermore, received signal strength (RSS) information and machine learning models have been utilized to detect tagged product movement type in retail store. Hauser et al [7] utilized RSS of RFID tag to obtain different pattern of tag movement so that it can help retail store to prevent shop theft. In addition, the RSS of RFID tag could be utilized as input for automated checkout systems by correctly detecting the items that leaving a store [8].

Nevertheless, there is no study about customer shopping behavior analysis (i.e., customer browsing activity) by utilizing RSS of RFID tag as an input for machine learning models. Therefore, in this study we proposed multilayer perceptron model to identify customer shopping behavior with the RSS value as an input. In addition, the statistical features were extracted from RSS of tags to be utilized as attribute for classifiers.

2. Methodology

In our study, the customer shopping behavior in retail store, i.e., customer activity when browsing products were investigated. Firstly, the RFID-enabled shelf was installed in the retail store. The RFID-enable shelf consists of single RFID reader and an antenna facing directly to the products where passive tags were attached on them. Secondly, receive signal strength (RSS) from the tags were collected for different customer behavior scenario. The statistical features were extracted from received signal strength (RSS) of tags. Finally, several machine learning algorithms were applied to distinguish different customer behavior on the tagged product. Experiments were carried out in a laboratory environment as a typical retail store scenario. Figure 1a shows the possible customer behavior that feasibly occur on products, such as the product is being browsed by customer and no behavior (customer does not have attention to the product). Figure 1b shows an example of tagged product is being browsed by customer. In our scenario, the time needed by the customers to browse each product is approximately less than 15 seconds and during each session, RSS of tag products were gathered for further analysis.



(a) **(b)**

Figure 1. Possible scenario of **(a)** customer shopping behavior and **(b)** example of customer is browsing the product.

We considered two type of tag reads, they are “no behavior” and “browsing” tag. “No behavior” represents situation where customer does not have any attention to the product, while “browsing” reveals the product is being browsed by customer. Single RFID reader ALR-9900+ and linear antennas ALR-9610-AL with 5.90 dbi Gain are utilized in this experiment [9]. In addition, UHF passive tags (model 9662, Alien H3) were attached to all products. We developed a gathering program based on Java programming language and installed in host computer connected to the reader. During each gathering

IOP Conf. Series: Materials Science and Engineering **803** (2020) 012022 doi:10.1088/1757-899X/803/1/012022

session, the reading data were stored as CSV file for further analysis. In total, 109 unique data readings were collected. The total number of data readings for "no behavior" and "browsing" are 62 and 47 respectively. For a gathering session, the received data or collection of tag event is expressed as follows:

$$X = \{\{ID, T_1, RSS_1, Ant_1, Class\}, \dots, \{ID, T_i, RSS_i, Ant_i, Class\}\} \tag{1}$$

where *X* denotes the data received by reader for tag *ID,* and *i* = *1,2, ..., n*. The parameter *n* represents the total number of tags occurs during a gathering session. Ant_i refers to the antenna which receives the tag data in the i_{th} time, while RSS_i denotes signal strength of the received data in the i_{th} time. The *Class* represent the class label of readings, where its value is either 0 (for "no behavior") or 1 (for "browsing").

Figure 2 shows example of RSS readings for a typical customer shopping behavior in retail store. For "no behavior", tag has relatively constant RSS, since the distance between the antenna and tagged product is fixed (Figure 2a). Figure 2b showed the product is being browsed by customer. The RSS decreases when tag stay away from antenna, achieving minimum when the tag is farthest from antenna or closest to the customer. When the product is returned to the original place, the RSS value increases and become constant again. As compared to the "no behavior" tag, the "browsing" tag tends to exhibit larger variance of RSS. However, different type of situation might arise in real case. Figure 2c showed the signal from "no behavior" or un-moved tag to antenna was blocked by the movement of other product. The RSS values decreases as other product move in between line of sight (LOS) of tag and antenna. Finally, figure 2d showed the tagged product is being browsed by customer but returned to the different place which is farther from antenna. The RSS value decreases as tag move away from antenna and generate lower RSS when it is placed farther from antenna. These conditions generate complex dataset of customer shopping behavior and appropriate attributes need to be extracted from time series dataset as input for classifiers.

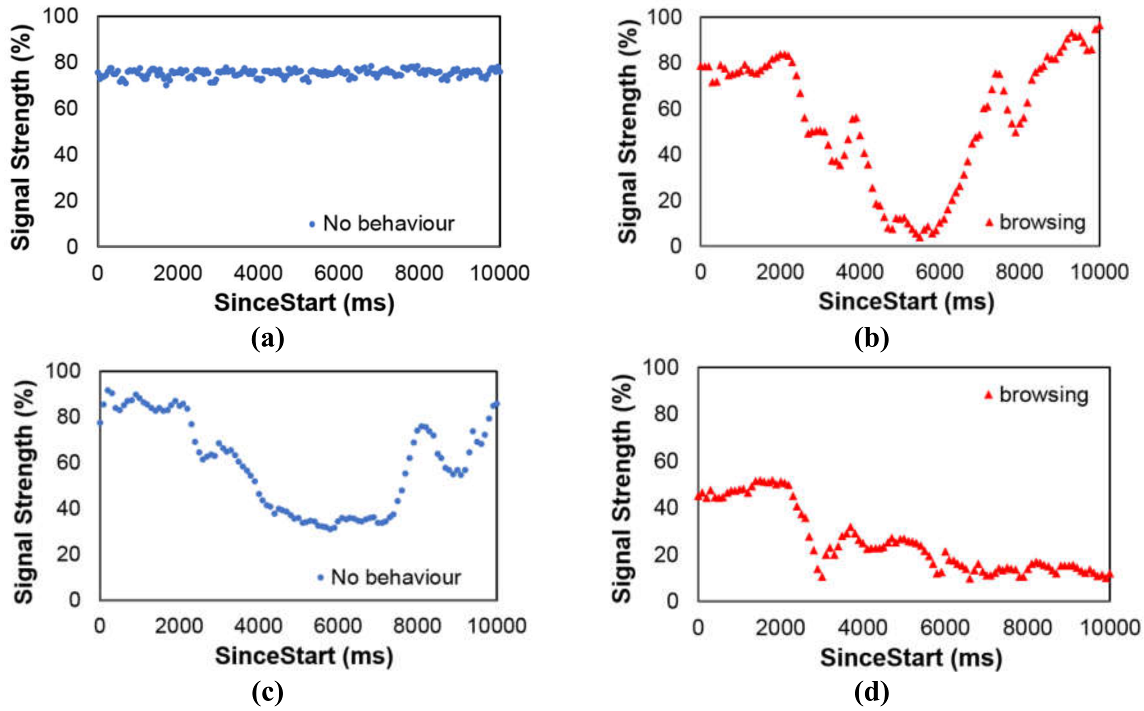


Figure 2. Example of RSS readings: **(a)** no behavior is conducted by customer to the product, **(b)** product is being browsed by customer, **(c)** signal of un-moved product is blocked by the movement of other product, and **(d)** product is being browsed and returned to different place by customer.

Since RSS information depends on the distance between the antenna and tag, closer tags generate larger RSS. Therefore, statistical information from RSS provided important information to differentiate

IOP Conf. Series: Materials Science and Engineering **803** (2020) 012022 doi:10.1088/1757-899X/803/1/012022

between moves and un-moved tags and have been utilized in previous studies [10,11]. Table 1 shows the proposed statistical features extracted from the RSS of single antenna.

Table 1. Statistical attributes extracted from a gathering session.

Attribute no	Attribute name	Description
1	min	Minimum signal strength during a gathering session.
2	max	Maximum signal strength during a gathering session.
3	mean	Average signal strength during a gathering session.
4	std	RSS standard deviation during a gathering session.
5	diff	Difference between highest and lowest signal strength during a gathering session.
6	median	Middle value signal strength during a gathering session.
7	kurtosis	Indicates if the signal strength distribution is heavy- or light-tailed, relative to normal.
8	skew	Distribution asymmetry of signal strength during a gathering session.
9	count	Total number of reads for the tag during a gathering session.

The preprocessing step needs to be conducted to convert tag readings into input matrix *X* and output vector *Y,* so that conventional machine learning model can learn and predict the outcome. Finally, given n different total tag occurrence in each gathering session, 9 (nine) total number of statistical features and m total unique tag readings data, the input X can be derived by creating the $[m \times 9]$ matrix

$$X = \begin{bmatrix} \min(RSS_{1,1}, \dots, RSS_{1,n}) & \max(RSS_{1,1}, \dots, RSS_{1,n}) & \dots & \text{count}(RSS_{1,1}, \dots, RSS_{1,n}) \\ \min(RSS_{2,1}, \dots, RSS_{2,n}) & \max(RSS_{2,1}, \dots, RSS_{2,n}) & \dots & \text{count}(RSS_{2,1}, \dots, RSS_{2,n}) \\ \vdots & \vdots & \vdots & \vdots \\ \min(RSS_{m-1,1}, \dots, RSS_{m-1,n}) & \max(RSS_{m-1,1}, \dots, RSS_{m-1,n}) & \dots & \text{count}(RSS_{m-1,1}, \dots, RSS_{m-1,n}) \\ \min(RSS_{m,1}, \dots, RSS_{m,n}) & \max(RSS_{m,1}, \dots, RSS_{m,n}) & \dots & \text{count}(RSS_{m,1}, \dots, RSS_{m,n}) \end{bmatrix} \tag{2}$$

and the $[m \times 1]$ output vector Y.

$$Y = \begin{bmatrix} Class_1 \\ Class_2 \\ \vdots \\ Class_{m-1} \\ Class_m \end{bmatrix} \tag{3}$$

In this study, we employed Multilayer Perceptron (MLP) model to distinguish the browsing activity of customer on the product. The MLP is a class of feedforward artificial neural network (ANN) with one input layer, one or more hidden layers, and one output layer. The backpropagation algorithm is utilized to train the MLP [12,13]. Net input was calculated by multiplying each input and its corresponding weight, and then summed. Each unit in the hidden layer took net input and then applied an activation function. Backpropagation compared the prediction result with the target class value and modified the weights for each training tuple to minimize mean squared error between prediction and target values. This process was iterated multiple times to produce optimal weights, providing optimal predictions for the test data. Furthermore, we employed feature selection method based on extremely randomized trees (Extra-Trees) algorithm [14] to remove irrelevant features and applied the final relevant attribute to the MLP, so that we expect to improve accuracy of proposed model.

The data preprocessing, feature selection and machine learning models were implemented in Python V3.6.6 and Scikit-learn V0.19.1 [15]. We used default parameters provided by Scikit-learn to simplify experiment process. 10-fold cross-validation were employed for all classification models.

3. Result and Discussion

Table 2 compares various model performances in terms of the percentage of accuracy, precision, recall, and f-score. The machine-learning models such as Logistic Regression (LR), Decision tree (DT), Naïve Bayes (NB), Random Forest (RF), AdaBoost, Support Vector Machine (SVM) are compared with the proposed model based on Multilayer Perceptron (MLP) to distinguish the customer shopping behavior. In this scenario, statistical features are extracted from RSS of tagged products and used as input attributes for classification models. In addition, the tree-based feature selection was applied only to proposed MLP model. The findings revealed that the proposed model outperformed other models by as much as 97.00%, 96.67%, 97.50%, and 96.57% for accuracy, precision, recall, and f-score, respectively.

Table 2. Performance of classification models on customer shopping behavior.

Method	Performance Evaluation (%)			
	Accuracy	Precision	Recall	F-score
LR	93.44	98.00	87.00	91.11
DT	93.18	95.00	93.50	92.74
NB	92.18	95.00	91.00	90.83
RF	94.18	95.71	93.50	93.34
AdaBoost	93.18	95.00	93.50	92.74
SVM	91.44	98.00	82.00	87.50
Proposed MLP	97.00	96.67	97.50	96.57

Furthermore, the impact of feature selection on the accuracy of classification models are presented in Figure 3. The result showed that by utilizing feature selection method for classifiers, provided higher accuracy as compared to utilizing all attributes as input, except for LR. Our result showed that by removing irrelevant features, it could improve accuracy of classifiers. Finally, by employing tree-based feature selection, the average of accuracy is improved as much as 1.019% compared to classifiers without feature selection method.

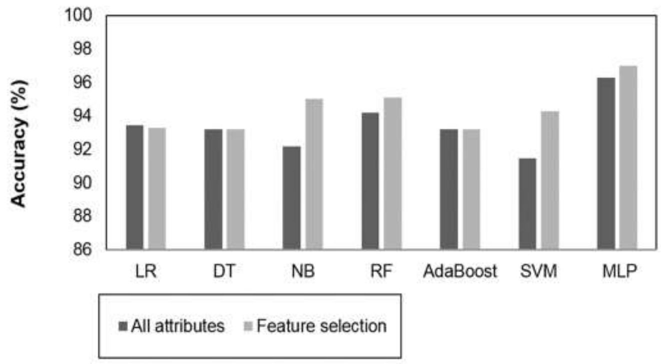


Figure 3. Impact of feature selection on classification accuracy.

The experimental results indicate that customer browsing activity on specific product can be detected by MLP model with high accuracy. By utilizing proposed model, the managers can better understand customer browsing pattern and what products customer interested in. Retail managers can use browsing and purchasing patterns for promotion as well as relevant product recommendations to the customers.

IOP Conf. Series: Materials Science and Engineering **803** (2020) 012022 doi:10.1088/1757-899X/803/1/012022

In addition, the managers might evaluate the product placement in retail store by removing the unpopular products from layout and promote other products that might boost their sales. Finally, as the interaction and quality of shopping increased, the customers' relationship with the retailer became stronger, which helped to lift the revenue of the business.

4. Conclusion and Future Works

This study proposed machine learning model to distinguish customer shopping behavior in retail store. The study utilized RFID-enabled shelf to track the movement of tagged products. The machine learning models were utilized to detect whether the tagged product was being browsed by customer or not. The RSS of tags were gathered, and statistical features were utilized as input for classification models. The proposed model based on Multilayer Perceptron (MLP) was compared with other classification models to detect the customer browsing activity on the products. The result revealed that the proposed model outperformed other models by as much as 97.00%, 96.67%, 97.50%, and 96.57% for accuracy, precision, recall, and f-score, respectively. Furthermore, the result showed that by employing tree-based feature selection, the average of accuracy was improved as much as 1.019% compared to classifiers without feature selection method.

Future study should consider more complex real situations, such as by considering other type of customer shopping behavior. Furthermore, employing different type of time series feature extraction and extending the comparison with other classification models, could be presented in the near future.

Acknowledgments

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and the Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (Grant number N0002301).

References

[1] Lee Y S, Yen S J, Tu G H and Hsieh M C 2004 Mining traveling and purchasing behaviors of customers in electronic commerce environment *Proc. of the IEEE Int. Conf. on e-Technology, e-Commerce and e-Service 2004 EEE '04 2004* pp 227-30

[2] Athauda T, Marin J C L, Lee J and Karmakar N C 2018 Robust Low-Cost Passive UHF RFID Based Smart Shopping Trolley *IEEE J. of Radio Frequency Identification* **2** no 3 pp 134-43

[3] Benes F, Stasa P, Svub J, Unucka J, Rhee J and Vojtech L 2017 Object Localization for Determining Customer's Behavior: Auto-ID Based Approach *Proc. of the 14th EAI Int. Conf. on Mobile and Ubiquitous Systems: Computing, Networking and Services* pp 468-77

[4] Syaekhoni M A, Alfian G and Kwon Y S 2017 Customer Purchasing Behavior Analysis as Alternatives for Supporting In-Store Green Marketing Decision-Making *Sustainability* **9** 2008

[5] Choi S H, Yang Y X, Yang B and Cheung H H 2015 Item-level RFID for enhancement of customer shopping experience in apparel retail *Computers in Industry* **71** pp 10-23

[6] Zhou Z, Shangguan L, Zheng X, Yang L and Liu Y 2017 Design and Implementation of an RFID-Based Customer Shopping Behavior Mining System *IEEE/ACM Transactions on Networking* **25** no 4 pp 2405-18

[7] Hauser M, Zügner D, Flath C M and Thiesse F 2015 Pushing the limits of RFID: Empowering RFID-based Electronic Article Surveillance with Data Analytics Techniques *Proc. of the 2015 Thirty Sixth Int. Conf. on Information Systems* Fort Worth TX USA 13–16 December 2015

[8] Hauser M, Günther S A, Flath C M and Thiesse F 2018 Towards Digital Transformation in Fashion Retailing: A Design-Oriented IS Research Study of Automated Checkout Systems *Bus. Inf. Syst. Eng.*

[9] Alien Technology 2019 *ALR-9900 Enterprise RFID Reader* http://www.alientechnology.com/wp-content/uploads/Alien-Technology-ALR-9900-Enterprise-RFID-Reader.pdf Retrieved May 30th 2019

[10] Ma H, Wang Y and Wang K 2018 Automatic detection of false positive RFID readings using

machine learning algorithms *Expert Syst. Appl.* **91** pp 442–51
[11] Alfian G, Syafrudin M, Yoon B and Rhee J 2019 False Positive RFID Detection Using Classification Models *Appl. Sci.* **9** 1154
[12] Han J, Kamber M and Pei J 2011 *Data Mining: Concepts and Techniques* 3rd ed M Kaufmann (MA USA: Publishers Burlington)
[13] Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back-propagating errors *Nature* **323** pp 533–6
[14] Geurts P, Ernst D and Wehenkel L 2006 Extremely randomized trees *Mach. Learn.* **63** 3
[15] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay É 2011 Scikit-learn: Machine Learning in Python *The J. of Machine Learning Research* **12** pp 2825-30